(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(72) Inventors: SEELEY, Albert, R.; 5 Brenda Lane, Burlington, MA 01803 (US). WILLIAMS, Douglas; 12 Abel Jones Place, Acton, MA 01720 (US). CHEN, Zhongyi; 3 Marie Circle, Burlington, MA 01803 (US). EDMONDSON, Robert; 35 Shute Street, Reading, MA 01867 (US).

(74) Agents: ROBINSON, Kermit et al.; Daly, Crowley & Mofford, LLP, Suite 101, 275 Turnpike Street, Canton, MA 02021 (US).

[Continued on next page]

(54) Title: METHOD OF TESTING A VOICE APPLICATION
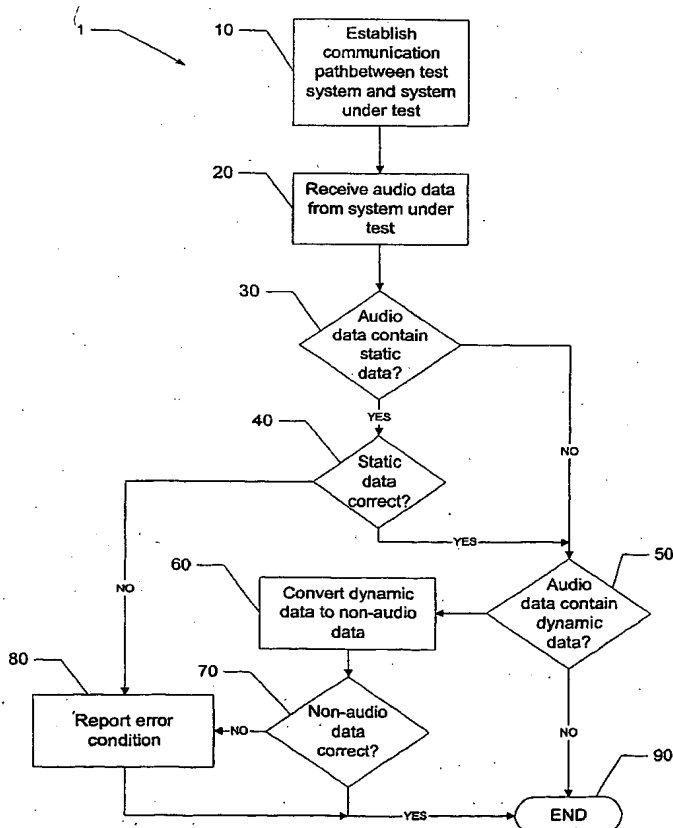
(57) Abstract: The present invention provides a method to automate the validation of dynamic data presented over telecommunications paths. The invention utilizes continuous speaker-independent speech recognition together with a process known generally as natural language recognition to reduce dynamic utterances to machine encoded text without requiring a prior training phase. Further, when configured by the end user to do so, the test system will convert common examples of dynamic speech, such as numbers, dates, times, and currency utterances into their usual textual representation.

WO 03/052739 A1

ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to the identity of the inventor (Rule 4.17(i)) for all designations*

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent*

(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for all designations*

Published:

— *with international search report*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

TITLE OF THE INVENTION

METHOD OF TESTING A VOICE APPLICATION

5   FIELD OF THE INVENTION

The present invention relates generally to voice application testing and more specifically to using automated speech recognition for web-based voice applications.

BACKGROUND OF THE INVENTION

10   Automated data provider systems are used to provide data such as stock quotes and bank balances to users over phone lines. The information provided by these automated systems typically comprises two parts. The first part of the information is known as static data. This can be, for example, a standard greeting or prompt, which may be the same for a number of users. The second part of the information is known as

15   dynamic data. For example, when providing a stock quote for a company the name of the company and the current stock price are dynamic data in the real world, because they change continuously as the users of the automated data provider systems make their selections and prices fluctuate.

In order to properly test such a system the automated data provider system needs

20   to be tested at two levels. One level of testing is to test the static data provided by the automated data provider. This can be accomplished, for example, by testing the voice prompts that guide the user through the menus, ensuring that the correct prompts are presented in the correct order. A second level of testing is to test that the dynamic data reported to the user is correct, for example, that the reported stock price is actually the

25   price for the named company at the time reported.

In existing test systems used to test automated data provider systems, speech data must be presented to the test system in a training phase prior to the testing phase, which prepares the system to recognize the same speech utterances when presented during the

30   testing phase. The recognition scheme is generally known as discrete speaker dependent

– 1 –

speech recognition. Thus, the system is limited to testing speech utterances presented to it *a priori*, and it is impractical to recognize dynamically changing utterances except where the set of all possible utterances is small.

One system that utilizes speech recognition as part of its provision of testing is the HAMMER IT™ test system available from Empirix Inc. of Waltham, MA. The HAMMER IT test system recognizes the responses from the system under test and verifies that the received responses are the same responses expected from the system under test. This test system works extremely well for recognizing static responses and for recognizing a limited number of dynamic responses which are known by the test system, however the HAMMER IT test system currently cannot test for a wide variety of dynamic responses which are unknown by the test system.

Another test system is available from Interactive Quality Systems (IQS) of Hopkins, Minnesota, which utilizes an alternative recognition scheme, namely, length of utterance, but is still limited to recognizing utterances presented to it *a priori*.

A possible alternative would be a semi-automated system, in which the dynamic portion of the utterance would be recorded and presented to a human operator for encoding. The dynamic portion of the utterance would be recorded and presented to a human operator for encoding in machine-readable characters.

In view of the above, it would be desirable to have a test system that tests the responses of automated data provider systems which presents both static data and dynamic data. It would be further desirable to have a test system which does not need to know beforehand the possible dynamic data.

SUMMARY OF THE INVENTION

The present invention provides a method to automate the validation of dynamic data (and static data) presented over telecommunications paths. The present invention utilizes continuous speaker-independent speech recognition together with a process known generally as natural language recognition to reduce dynamic utterances to machine encoded text without requiring a prior training phase. Further, when configured by the end user to do so, the test system will convert common examples of dynamic

speech, such as numbers, dates, times, and currency utterances into their usual textual representation. For instance, the test system will convert the utterance "four hundred fifty four dollars and twenty nine cents" into the more usual representation of "454.29". This will eliminate the limitation that all tested utterances need to be known by the test
5   system in advance of the test.

By converting the dynamic utterances to machine encoded text, the invention facilitates automated validation of the data so converted, by allowing use of the converted data as input into an automated system which can independently access and validate the data.

10   Additionally, it is an object of the present invention to utilize Automated Speech Recognition (ASR) to perform several functions. These functions which utilize ASR include monitoring of Interactive Voice Response (IVR) applications, testing web-based voice applications, and using ASR in a hosted service environment. A command set is implemented to provide a programming interface between the testing/monitoring systems
15   to the ASR functionality.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will be better understood by reference to the following more detailed description and accompanying drawings in which:
20   Fig. 1 is a flow chart of the presently disclosed method.

DETAILED DESCRIPTION

Proper testing of an automated data provider system requires the ability of the automated system performing the test to provide two functions. One function is the
25   testing of static audio data received from the system under test. The audio data is received and processed and speech recognition is performed. The static portion of the utterance is validated against the expectations for the current state of the system under test. A second function of the test system is to provide a conversion from the verbal report of the data (dynamic data) by the system under test into a textual representation.
30   The textual representation, typically in the form of machine encoded characters, is then

- 3 -

used as an input into an automated system which can independently access the data in question and validate the accuracy of the response. For example, in the case of a stock quotation, accessing the stock exchange database and comparing the results of the access with the textual representation of the dynamic data verify the textual representation of the

5   dynamic data.

One advantage of the present invention is that it directly reduces arbitrary dynamic utterances presented over telecommunications devices, such as dollar amounts, times, account numbers, and so on, into machine encoded character representations suitable for input into an automated independent validation system, without intermediate

10   human intervention. Another advantage afforded by the present invention is that it eliminates the limitation imposed on known test systems that all possible tested utterances are known in advance of the test.

In the presently disclosed invention, the result of the testing of data from an automated data provider system will be one or more of the following three results. First,

15   a text string of the recognized words, for example, "Enter|pin|number|". Second, natural language "understanding" of the speech clip, so that, for example, "five hundred twelve dollars and thirty five cents" would be recognized as $512.35. Third a tag, which is a user defined name for a recognized utterance.

In addition, the presently disclosed system is able to perform speaker independent

20   recognition, so that creating a vocabulary of static utterances is not necessary.

A flow chart of the presently disclosed method is depicted in Figure 1. The rectangular elements are herein denoted "processing blocks" and represent computer software instructions or groups of instructions. The diamond shaped elements, are herein denoted "decision blocks," represent computer software instructions, or groups of

25   instructions which affect the execution of the computer software instructions represented by the processing blocks.

Alternatively, the processing and decision blocks represent steps performed by functionally equivalent circuits such as a digital signal processor circuit or an application specific integrated circuit (ASIC). The flow diagrams do not depict the syntax of any

30   particular programming language. Rather, the flow diagrams illustrate the functional

– 4 –

information one of ordinary skill in the art requires to fabricate circuits or to generate computer software to perform the processing required in accordance with the present invention. It should be noted that many routine program elements, such as initialization of loops and variables and the use of temporary variables are not shown. It will be

5   appreciated by those of ordinary skill in the art that unless otherwise indicated herein, the particular sequence of steps described is illustrative only and can be varied without departing from the spirit of the invention. Thus, unless otherwise stated the steps described below are unordered meaning that, when possible, the steps can be performed in any convenient or desirable order.

10   The first step 10 of the process is to establish a communications path between the test system and the system under test. This communications path may be a telephone connection, a wireless or cellular connection, a network or Internet connection or other types of connections as would be known by someone of reasonable skill in the art.

Step 20 comprises receiving audio data from the system under test by the test

15   system through the communication path established in step 10. This received audio data may include static data, dynamic data or a combination of static and dynamic data. As an example, the list below contains the possible instances of audio data to be received from the system under test.

"This is the MegaMaximum bank"

20   "If you need assistance at any time, just say Help"

"Please enter or say your account number"

"Please enter or say your pin number"

"Your current balance is <dollars>"

"We're sorry, your account number or pin were not recognized. Please try again."

25   "An associate will be with you shortly."

Once the audio data is received, at step 30 a determination is made as to whether the audio data contains static data. In the case where the audio data comprises "This is the MegaMaximum bank", the entire data is static data. In the case wherein the audio

- 5 -

data received is "Your current balance is <dollars>" a combination of static data ("Your current balance is") and dynamic data ("<dollars>") has been received.

At step 40, a determination is made as to whether the static data is correct. If the static data corresponds to the expected data the static data is deemed correct, then step 50 is executed. If the static data does not correspond to the expected data the static data is deemed incorrect, then an error condition is indicated as shown in step 90.

Following step 30 if no static data has been received, or step 40 if the static data received is correct, step 50 is executed. At step 50 a determination is made as to whether the received audio data contains dynamic data. If no dynamic data has been received, then step 80 is executed, and the process ends. If dynamic data has been received as part of the received audio data, then step 60 is executed.

Step 60 converts the dynamic data to non-audio data. This non-audio data can be, for example, a textual format such as machine encoded text. Other formats could also be used. Following the conversion of dynamic data to non-audio data, step 70 is executed.

Step 70 determines whether the non-audio data is correct. The non-audio data could be a stock price, a dollar amount, or the like. This non-audio data typically is compared to a database which contains the correct data. If the non-audio data was correct, then step 80 is executed and the process ends. If the non-audio data was not correct then step 90 is executed wherein an error condition is reported.

Referring back to the example dynamic data phrase "Your current balance is <dollars>" which contains the dynamic data, the user would construct a grammar to inform the recognizer of the expected utterances and their interpretation, so that, for example, the "<dollars>" slot would be interpreted as a monetary amount ("$512.00") rather than a string of words ("five|hundred|twelve|dollars|and|zero|cents|"). The grammar could also assign tags (names) to each utterance, which the recognizer would return along with the text and/or interpretation. For the simpler applications, this would provide a solution conceptually similar to how prompt recognition is typically performed. The grammar would correspond to the vocabulary, and the tag would be a symbolic version of the clip number received as a recognition result.

- 6 -

Grammars are constructed as text files, with a GUI (Graphical User Interface) interface to ease the user through the arcane syntax. A pseudo-grammar might look as follows:

5   <phrase1> = (this is the megamaximum bank) {greeting}

<phrase2> = (if you need assistance just say help) {help_prompt}

<phrase3> = (please enter or say your account number) {account}

<phrase4> = (please enter or say your pin number) {pin}

<dollars> = [NUMBER]

10   <phrase5> = (your current balance is <dollars>{amount}){balance}

...

In the above examples, the elements inside the curly braces ("greeting", "help_prompt", "amount", etc.) comprise the tags which are returned if their corresponding phrase were recognized.

When running the script, as each prompt is presented by the system under test, the prompt is sent off to be recognized, and a string, tag, and understanding, if any, are returned as the result. The script compares the returned string against the expected string, or simply checks the tag to see if it is the expected one. For the phrase "your current balance is <dollars>{amount}{balance}" above, the script compares only the first four words (static data – "your current balance is"), and compares the dollar amount (dynamic data - <dollars>) to the expected value as a separate operation.

To implement this, the following is required. A utility to enroll "MegaMaximum" into the speech recognizer's vocabulary. Another utility to set up a grammar. A command to connect the running script with the created grammar. Another command to compare strings and substrings on a word-by-word basis (rather than the character basis of most string utilities). A command to retrieve the "next slot" from the returned result, such as the <dollars> item from phrase number five. Another command to detect speech and "barge in" with the request for help. Another command to send the utterance to the new recognizer and obtain the result structure. In a particular embodiment the result structure would nominally include the status (recognized, failed), the tag (name) of the

- 7 -

utterance, a probability score (0-100, with 100=best), and the text rendition of the utterance. If language understanding were performed, such as the translation of numeral names into currency, the recognized sub-portions would be included in the result structure as well.

5　　　　As described above, the presently disclosed invention performs recognition on larger and more varied utterances than currently available systems. Further, the presently disclosed invention handles dynamic data seamlessly with static data.

One application involves the use of ASR for monitoring IVR applications. In this application test telephone calls are generated by a test system to an IVR and the speech

10　　responses are actively monitored. Prompts provided by the system under test are captured and analyzed for performance and accuracy.

One method utilized to transform human-readable text into speech is known as Text-To-Speech (TTS). TTS is often used in conjunction with Automated Speech Recognition (ASR) systems to render prompts with embedded dynamic speech elements.

15　　TTS may be used to convert either of a literal text string or text contained in a file.

Other applications involving the use of ASR are also provided. ASR is used to develop testing and monitoring solutions for web-based voice applications built on defined technologies. These technologies include standards for voice data such as Voice XML and Speech Application Language Tags (SALT). ASR may also be used as a core

20　　component of hosted services that provide both voice application load testing and voice application monitoring.

In a particular embodiment the programming interface to the ASR functionality from a test system comprises the following commands: AsrEnableSpeech, AsrDisableSpeech, AsrRecognize, AsrRecognizeFile, AsrRecognizePartial,

25　　AsrGetResults, AsrGetAnswer, AsrGetSlot, AsrSetParameter, and AsrGet Parameter.

A method to automate the validation of dynamic data presented over telecommunications paths has been described. The invention utilizes continuous speaker-independent speech recognition together with a process known generally as natural language recognition to reduce dynamic utterances to machine encoded text without

30　　requiring a prior training phase. Further, when configured by the end user to do so, the

- 8 -

test system will convert common examples of dynamic speech, such as numbers, dates, times, and currency utterances into their usual textual representation.

Having described preferred embodiments of the invention it will now become apparent to those of ordinary skill in the art that other embodiments incorporating these
5   concepts may be used. Additionally, the software included as part of the invention may be embodied in a computer program product that includes a computer useable medium. For example, such a computer usable medium can include a readable memory device, such as a hard drive device, a CD-ROM, a DVD-ROM, or a computer diskette, having computer readable program code segments stored thereon. The computer readable
10 medium can also include a communications link, either optical, wired, or wireless, having program code segments carried thereon as digital or analog signals. Accordingly, it is submitted that that the invention should not be limited to the described embodiments but rather should be limited only by the spirit and scope of the appended claims. All publications and references cited herein are expressly incorporated herein by reference in
15 their entirety.

− 9 −

CLAIMS

We claim:

1. A method comprising:

establishing a communications path between a test system and a system under test (SUT);

receiving by said test system, audio data from said SUT;

determining whether said audio data contains static data, and when said audio data contains static data, verifying the correctness of said static data;

determining whether said audio data contains dynamic data, and when said audio data does contain dynamic data, converting said dynamic data to non-audio data and verifying the correctness of said non-audio data; and

reporting an error condition when at least one of said non-audio data and said static data is not correct.

2. The method of claim 1 wherein said non-audio data comprises text.

3. The method of claim 2 wherein said text comprises machine-encoded characters.

4. The method of claim 1 wherein said verifying the correctness of said non-audio data comprises independently acquiring data and comparing the independently acquired data to said non-audio data.

5. The method of claim 1 wherein said converting comprises utilizing natural language recognition.

6. The method of claim 1 wherein said converting includes converting common examples of dynamic data to their usual textual representation.

- 10 -

7.   The method of claim 6 wherein said common examples of dynamic data includes numbers, dates, times and currency.

8.   The method of claim 1 wherein said converting includes providing a tag for
5   identifying said non-audio data.

9.   A computer program product, disposed on a computer readable medium, the computer program product including instructions for causing a processor to:

        establish a communications path between a test system and a system under test
10   (SUT);

        receive audio data from said SUT;

        determine whether said audio data contains static data, and when said audio data contains static data, verify the correctness of said static data;

        determine whether said audio data contains dynamic data, and when said audio
15   data does contain dynamic data, convert said dynamic data to non-audio data and verify the correctness of said non-audio data; and

        report an error condition when at least one of said non-audio data and said static data is not correct.

20   10.   The computer program product of claim 9 wherein said non-audio data comprises text.

11.   The computer program product of claim 10 wherein said text comprises machine-encoded characters.

25

12.   The computer program product of claim 9 wherein said instructions for causing a processor to verify the correctness of said non-audio data comprises instructions for causing the processor to independently acquire data and compare the independently acquired data to said non-audio data.

30

- 11 -

13.   · The computer program product of claim 9 wherein said instructions for causing a processor to convert said dynamic data to non-audio data comprises utilizing natural language recognition.

5   14.   The computer program product of claim 9 wherein said instructions for causing a processor to convert said dynamic data to non-audio data includes instructions for causing the processor to convert common examples of dynamic data to their usual textual representation.

10   15.   The computer program product of claim 14 wherein said common examples of dynamic data includes numbers, dates, times and currency.

16.   The computer program product of claim 9 wherein said instructions for causing a processor to convert said dynamic data to non-audio data includes instructions for
15   causing the processor to provide a tag for identifying said non-audio data.

- 12 -

1/1
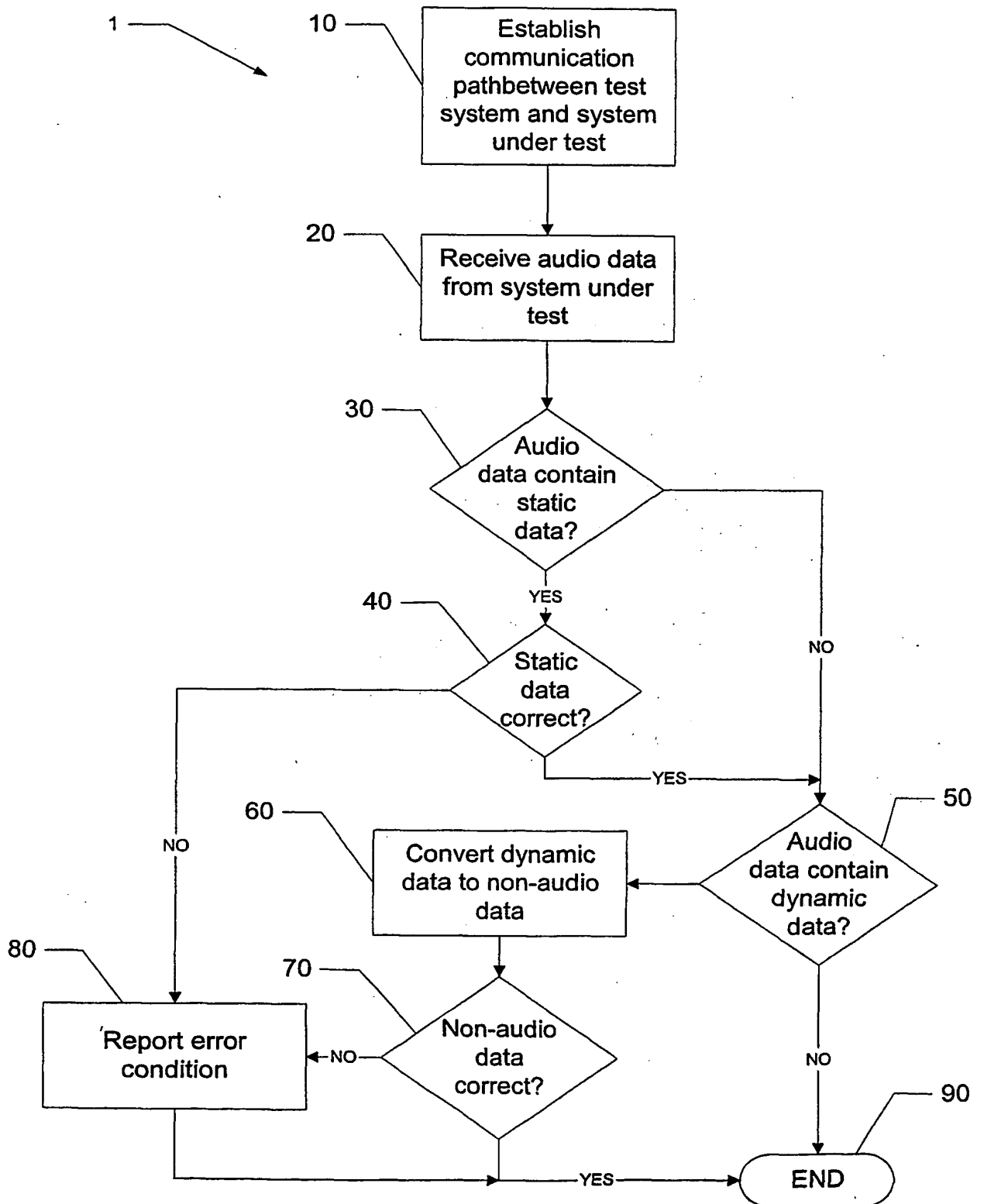
FIGURE 1

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7   G10L15/26   H04M3/493

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7   G10L   H04M

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	US 2002/138261 A1 (ZIEGELMILLER) 26 September 2002 (2002-09-26) the whole document	1-16
X	US 6 321 198 B1 (HANK ET AL) 20 November 2001 (2001-11-20) abstract; figures 4,5,12 column 2, line 35-50	1-16
A	SATURNINO LUZ: "STATE-OF-THE-ART SURVEY OF EXISTING DIALOGUE MANAGEMENT TOOLS" ESPRIT LONG TERM RESEARCH CONCERTED ACTION N° 24823, 22 June 2000 (2000-06-22), XP002239586 NIS Lab /luz, Odense University, DK the whole document	1-16

☐ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
28 April 2003	13/05/2003

| Name and mailing address of the ISA
European Patent Office, P.B. 5818 Patentlaan 2
NL – 2280 HV Rijswijk
Tel. (+31–70) 340–2040, Tx. 31 651 epo nl,
Fax: (+31–70) 340–3016 | Authorized officer

Quélavoine, R |

Patent document cited in search report		Publication date		Patent family member(s)		Publication date
US 2002138261	A1	26-09-2002	WO	02077976	A1	03-10-2002
US 6321198	B1	20-11-2001	CA	2365743	A1	31-08-2000
			EP	1155371	A1	21-11-2001
			JP	2002543445	T	17-12-2002
			WO	0051016	A1	31-08-2000

THIS PAGE BLANK (USPTO)